The Vanderbilt University CENTER FOR DIGITAL humanities

Topic Modeling Race-Talk in Nashville's White Christian Churches

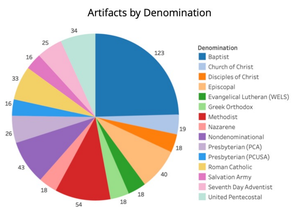
Debbie Brubaker – Ph.D. Candidate, Vanderbilt University, Graduate Department of Religion

Overview

In 2020 and 2021, the violence of anti-Black racism and white supremacy was on display and widely discussed across the United States. This project uses topic modeling techniques to analyze digital artifacts produced by white churches during this time. I ask: **when and how did white Christian churches in Nashville, TN (digitally) discuss race or racism between May 2020 and February 2021?**

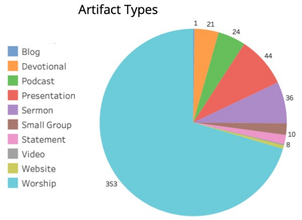
This project is part of a larger qualitative and theological study of white-majority Christian churches in Nashville, TN, and contributes to its transdisciplinary design by evidencing how digital humanities methods aid the study of religious phenomena.

Dataset



The dataset for this project includes 501 total digital artifacts. The artifacts were produced by a group of 22 white-majority Christian churches based in Nashville, TN. The chart to the left demonstrates the number of artifacts according to the denominational affiliation of their community of origin.

The digital artifacts included in this project are transcripts of audio, visual, and textual content produced by these churches and published online. The artifacts represent diverse social practices and religious rituals, including worship services, podcasts, organizational statements, and religious education sessions.



Methods

- Data Collection
- Transcription
- Corpus Creation and Preprocessing
- Topic Modeling
- More Topic Modeling
- Conversion to Tabular Data
- Data Analysis and Visualization

Qualitative analysis throughout

Word cloud of document with highest probability of "race/racism" topic (visualization made with Voyant Tools)

Data collection was conducted through web scraping. Transcriptions were generated using **Nvivo** Transcription and were left unedited.

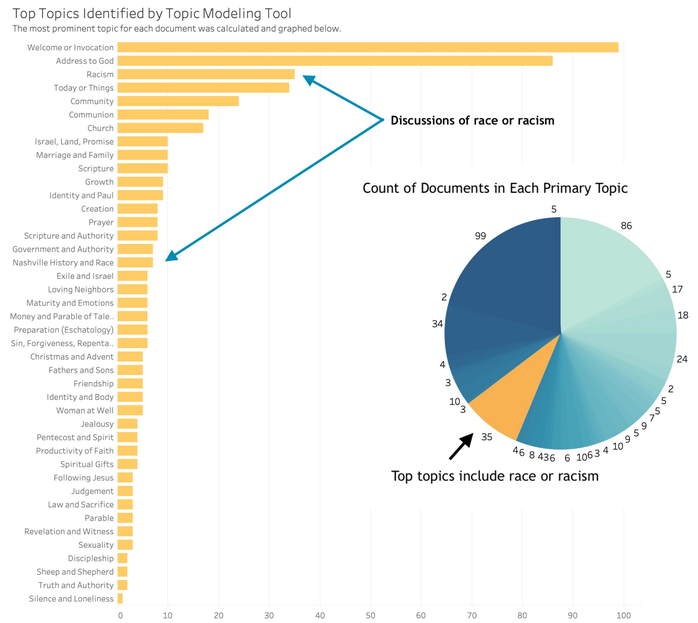
The **tm package in R** was used to create and preprocess the corpus of documents.

Topic modeling analysis was conducted iteratively, modeling for topic counts between 8 and 200. I experimented by using two separate tools (**topicmodels** in R and the **Topic Modeling Tool**) at multiple intervals to identify coherent topics.

Output from the topic modeling analysis was then converted to tabular data and combined with additional data about the document source. Visualizations were created using this dataset in **Tableau**, while **Voyant Tools** enabled visualizations of the corpus files.

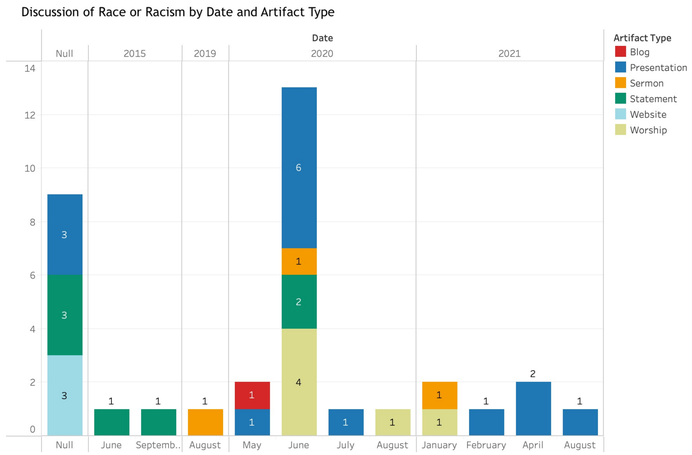
Findings

Both topic modeling tools produced coherent topics connected to themes of race, racism, or the history of racism in Nashville and the United States. The tools identify the most prominent topics in each artifact of the corpus, thereby suggesting key documents within a large corpus to be targeted for further analysis and close reading.



Race and racism topics are **most** frequent in late May and June 2020. These topics also appear in January and February of 2021, but do **not** appear between October-November 2020.

Topics about race and racism **most** frequently appear in special presentations, website content, or text-based institutional statements. Topics about race appear **less** frequently in worship service artifacts.



Next Steps

Targeted Coding: Further analysis of individual documents is needed to evaluate and refine topic identification.

Targeted Close Reading: Topic modeling provides a valuable "distant reading" of a mid-to-large size corpus. Yet, a close reading of specific artifacts is necessary for an accurate exploration of artifact content.

Network and Timeline Analysis: Assessment of relationships between topics and the timeline of topic appearance will clarify artifact content and context.

References

- Awati, Kailash. "A Gentle Introduction to Topic Modeling Using R." *Eight to Late* (blog), September 29, 2015. https://eight2late.wordpress.com/2015/09/29/a-gentle-introduction-to-topic-modeling-using-r/.
- Butler, Anthea. *White Evangelical Racism: The Politics of Morality in America*. Chapel Hill, NC: The University of North Carolina Press, 2021.
- Enderle, Jonathan Scott, Arun Balagopalan, Xiaojing Li, and David Newman. "Senderle/Topic-Modeling-Tool: First Stable Release." April, 2017. https://github.com/senderle/topic-modeling-tool.
- Graham, Shawn, Scott Weingart, and Ian Milligan, "Getting Started with Topic Modeling and MALLET," *Programming Historian*, September 2, 2012. https://programminghistorian.org/en/lessons/topic-modeling-and-mallet.
- Grün, Bettina, and Kurt Hornik. "**Topicmodels**: An *R* Package for Fitting Topic Models." *Journal of Statistical Software* 40, no. 13 (2011). https://doi.org/10.18637/jss.v040.i13.
- Guldi, Jo. "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora." *Journal of Cultural Analytics* 3, no. 1 (December 20, 2018).
- Jones, Robert P. "Whiteness after the Election: Where Do We Go from Here?" Vanderbilt University (Virtual), November 12, 2020. https://www.facebook.com/vandydiv/videos/299419411156595.
- Posner, Miriam, and Andy Wallace. "Very Basic Strategies for Interpreting Results from the Topic Modeling Tool." *Miriam Posner's Blog* (blog), October 29, 2012. https://miriamposner.com/blog/very-basic-strategies-for-interpreting-results-from-the-topic-modeling-tool/.
- Saxton, Micah. "Topic Modeling Best Practices." Accessed April 14, 2022. https://msaxton.github.io/topic-model-best-practices/lit_review.html.
- Underwood, Ted. "Topic Modeling Made Just Simple Enough." *The Stone and the Shell* (blog), April 7, 2012. https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/.

Acknowledgements

This project was funded by an Andrew W. Mellon Fellowship for the Digital Humanities. Critical support and engagement at various stages of the project came from the directors of The Center for Digital Humanities at Vanderbilt University, Lynn Ramey and Madeleine Casad, and the 2021-2022 fellows. I would also like to thank the participants of the Text Analysis Working Group, especially Mark Schoenfield and Steve Baskauf, for their engagement and technical support throughout the project.